Drug-repurposing and discovery in the era of "big" real-world data: how the incorporation of observational data, genetics, and other -omic technologies can move us forward

Megan M. Shuey
Division of Genetic Medicine, Department of Medicine, Vanderbilt Genetics Institute
Vanderbilt University Medical Center
2525 West End Ave. Ste 700, Nashville, TN, 37203, USA
Email: megan.m.shuey.1@vumc.org

Jacklyn N. Hellwege
Division of Genetic Medicine, Department of Medicine, Vanderbilt Genetics Institute
Vanderbilt University Medical Center
2525 West End Ave. Ste 700, Nashville, TN, 37203, USA
Email: jacklyn.hellwege@vumc.org

Nikhil Khankari
Division of Genetic Medicine, Department of Medicine, Vanderbilt Genetics Institute
Vanderbilt University Medical Center
2525 West End Ave. Ste 700, Nashville, TN, 37203, USA
Email: nikhil.khankari@vumc.org

Marijana Vujkovic
Division of Translational Medicine and Human Genetics
University of Pennsylvania Perelman School of Medicine
3400 Civic Center Blvd, Philadelphia, PA 19104, USA
Email: vujkovic@pennmedicine.upenn.edu

Todd L. Edwards
Division of Epidemiology, Vanderbilt Genetics Institute, Vanderbilt University Medical Center
2525 West End Ave. Ste 600, Nashville, TN, 37203, USA
Email: todd.l.edwards@vumc.org

This PSB 2024 session discusses the many broad biological, computational, and statistical approaches currently being used for therapeutic drug target identification and repurposing of existing treatments. Drug repurposing efforts have the potential to dramatically improve the treatment landscape by more rapidly identifying drug targets and alternative strategies for untreated or poorly managed diseases. The overarching theme for this session is the use and integration of real-world data to identify drug-disease pairs with potential therapeutic use. These drug-disease pairs may be identified through genomic, proteomic, biomarkers, protein interaction analyses, electronic health records, and chemical profiling. Taken together, this session combines novel applications of methods and innovative modeling strategies with diverse real-world data to suggest new pharmaceutical treatments for human diseases.

Keywords: drug repurposing, drug repositioning, -omics, machine learning, genetics, translational science

1. Drug development and repurposing

Drug discovery and development is a long and high-risk process with cumulative annual costs approaching $1 billion US dollars (Hinkson et al. 2020),(Wouters et al. 2020), where over 85% of drug candidates will fail prior to completing clinical trials (Nielsch et al. 2016). Drug repurposing or repositioning of existing medications for new therapeutic uses can substantially reduce costs, time, and effort while providing additional treatment options to patients.

The increasing availability of large-scale electronical medical record (EMR) data, in combination with genomic, proteomic, molecular, and other biomedical data is enabling more cost-effective investigations of treatment response, adverse event profiling, and novel target identification. The use of "real-world data" presents a promising solution with the potential to dramatically reduce development time and cost. Furthermore, policy makers in the US and other countries are increasingly open to considering alternative sources of evidence beyond clinical trials in their decision-making processes. For instance, the 21st Century Cures Act encourages the use of real-world data to generate evidence of product effectiveness to help support approval of new indications for existing drugs (Dagenais et al. 2022).

Traditionally these sources encompass data from hospital or population-based health records, third-party health insurance claims, registries, and health surveys (Administration 2018). These data types are increasingly linked to novel types of biomedical data, such as genomic (or other "omic") data from large biobanks, biopsies, pathology tests, diagnostic imaging, and information related to social determinants of health (SDoH). Initiatives focused on data-driven drug repurposing, leveraging the expansive resources available within real-world data repositories have the potential to improve the efficiency of identifying potential treatments, while simultaneously reducing possible risks associated with drug development.

Computational approaches, such as machine learning, offer a powerful avenue to address specific challenges in drug development by harnessing the wealth of multi-dimensional data from various sources. For example, prior research has demonstrated the ability of machine learning algorithms to scan compound libraries to optimize the design of small molecules and evaluate molecular docking to estimate drug-target interactions, and use this to find repurposing targets for viral infections and cancer (Kumar et al. 2015; Mirza et al. 2016; Wang et al. 2017).

Phenotype-first approaches used in conjunction with machine learning are yet another example of identifying targets for drug repurposing. This methodology identifies a set of optimal treatment modalities using medical record history which has demonstrated increased efficacy of clinical conditions. This approach capitalizes on the growing availability of EMR data to evaluate acute and long-term therapeutic response based on individual-level, real-world clinical data. These deep-

learning computational approaches can be coupled with genetics and advanced -omics data to elucidate the underlying mechanisms of disease. When combined with available drug-target datasets, this information can facilitate the identification of alternative treatment strategies (Allen et al. 2015; MacEachern and Forkert 2021; Wang et al. 2021; Xu et al. 2022).

Similarly, novel computational approaches that leverage genomic and transcriptomic methodologies, including but not limited to genome-wide association studies, genetically predicted gene expression analysis, and Mendelian randomization have the potential to identify and estimate the effect of drug repurposing on reducing risk of disease. These approaches are particularly appealing, given that drugs with genetic evidence from disease association studies have a two-fold higher likelihood of successfully reaching the market (Nelson et al. 2015; King et al. 2019).

The research team encompassing this panel has experience in developing such computational pipelines for identification of potential drug candidates for repurposing in diabetes treatment (Khankari et al. 2022; Shuey et al. 2022). This approach specifically uses a transcriptome-driven drug screening approach to identify candidate therapeutics. Subsequently, it validates these candidates through a two-step process by: 1) generating real-world evidence for drug efficacy using a self-controlled case series study design using large EMR datasets and quantifying changes in disease-associated biomarkers before and after treatment with identified candidates, and 2) generate genetic evidence for drug target efficacy for disease using the Mendelian randomization framework. We encourage participation in this series by other researchers who are involved in the development of strategies to aid in the identification and evaluation of drug repurposing opportunities.

2. Session contents

Here we describe briefly studies which will be presented during the session.

2.1. *List of topics captured in this session*

Our session includes presentations on the following diverse topics related to drug repurposing and discovery:

- Modeling of outcome risk based on medication exposure using propensity score matching
- Improved techniques for target identification from sequencing data
- Machine learning modeling of disease and protein interaction networks (???)
- High-throughput functional screening assays

2.2. *Systematic Estimation of Treatment Effect on Hospitalization Risk as a Drug Repurposing Screening Method*

In this manuscript, *Georgantas et al.* propose a simple, pragmatic screening approach for drug repurposing using real-world data. They incorporated time-to-event and propensity score matching with observational data from the UK Biobank to evaluate the roles of thousands of drug-disease pairs on hospitalization risk. This elegant use of high-dimensional real-world data suggests numerous repurposing opportunities for existing, commonly prescribed medications.

2.3. *Transcript-aware analysis of rare predicted loss-of-function variants in the UK Biobank elucidate new isoform-trait associations*

As whole exome and genome sequencing becomes more widely accessible, the ability to synthesize these results into meaningful discoveries is essential. Traditional burden testing approaches assume that all variants in a given gene have similar effects on gene function and fails to consider isoforms where this assumption is often violated. *Hoffing et al.* demonstrates how using transcript-specific annotations (rather than collapsed gene-based evaluations) to classify rare predicted loss-of-function (pLOF) mutations can dramatically impact effect estimates for rare variant association analyses. Their work links such pLOFs to tissue specificity, quantitative endophenotypes, and disease outcomes and has a distinct outcome for improving the outputs of such large-scale sequencing data for drug target identification. The results of this study have the potential to improve accuracy of rare variant-disease association studies that often serve to identify novel drug targets.

2.4. *Formulating new drug repurposing hypotheses using disease-specific hypergraphs*

In *Jain et al.*, the authors use disease-specific hypergraphs in which hyperedges of various lengths encode biological pathways to generate new repurposing targets which may be overlooked by classic knowledge graphs. These low-dimensional representations of drug-to-gene pathways are filtered to existing therapeutic approaches for Alzheimer's Disease and then evaluated using the multiscale interactome (MSI). Further, the seven targets not represented in MSI were evaluated by literature review, with many of these candidates having demonstrable impacts on brain development or disease processing that support a relationship with Alzheimer's Disease.

2.5. *Combined kinome inhibition states are predictive of cancer cell line sensitivity to kinase inhibitor combination therapies*

Kinase inhibitors are a staple in clinical oncology; however, monotherapy may lead to resistance in part due to compensation by other members of the kinase network or kinome. Combinatorial therapies have been suggested to combat this resistance. However, determining the best combination of kinase inhibitors is essential. To this end, *Joisa et al.* developed a high-throughput platform for evaluating combinatorial effects of multiple kinase inhibitors. By leveraging heterogenous data for the prediction of potential drug combination targets the authors identified the combination of MEK and PI3K inhibitors (Trametinib/Omipalisib). Their results are supported by this particular combination of inhibitors entering a recent phase 1 clinical trial which suggesting the potential for this method to identify other combinatorial therapies.

2.6. *The Human Protein Structure Targetome*

Ovanessians et al. utilized structure-based modeling of proteins for more than 20,000 human proteins curated from various protein databases to build a human "targetome". This approach was developed to prioritize protein-ligand pairs and accounts for the complexities of both protein structure and binding site affinities to prioritize drug targeting. The potential of this pipeline and strategies like this have the potential to advance drug design and development efforts by not only

prioritizing candidates but informing various considerations in the drug development pipeline including competitive binding estimation.

2.7. *Modeling Path Importance for Effective Alzheimer's Disease Drug Repurposing*

The final manuscript in this session by *Xiang et al.* presents a modeling schema focused on building a large-scale protein-protein interaction network from various data sources. Their approach incorporated both available data about protein-protein interactions and existing drug-target interactions to develop a rich data resource for prioritization of biological systems, e.g. networks and pathways. Their models captured a network's rich topology and challenges the assumption that paths of equal length have equivalent importance in biological systems. Results were further supported by the prioritization of several drug candidates that are supported by previous publications and insurance claims data.

3. Conclusion

The authors in this session present six diverse papers that discuss methodologic improvements to guide potential drug discovery and repurposing. The session expands upon the application of commonly used techniques like improving prediction of loss-of-function mutations for target identification as well as modeling strategies using genetic data to evaluate medication exposure and outcomes. There is also a special emphasis on using machine learning techniques and available datasets to identify drug targets by considering disease, protein, and kinase interactions. We anticipate that these studies, results, and associated techniques can advance disease-specific target evaluation and drug repurposing.

Acknowledgements

JHN and MMS were supported by K12-HD043483. MV, TLE, and NK are supported by R01-DK134575.

References

Administration USFaD. 2018. Framework for FDA's Real World Evidence Program.

Allen BK, Mehta S, Ember SW, Schonbrunn E, Ayad N, Schurer SC. 2015. Large-Scale Computational Screening Identifies First in Class Multitarget Inhibitor of EGFR Kinase and BRD4. *Sci Rep* **5**: 16924.

Dagenais S, Russo L, Madsen A, Webster J, Becnel L. 2022. Use of Real-World Evidence to Drive Drug Development Strategy and Inform Clinical Trial Design. *Clin Pharmacol Ther* **111**: 77-89.

Hinkson IV, Madej B, Stahlberg EA. 2020. Accelerating Therapeutics for Opportunities in Medicine: A Paradigm Shift in Drug Discovery. *Front Pharmacol* **11**: 770.

Khankari NK, Keaton JM, Walker VM, Lee KM, Shuey MM, Clarke SL, Heberer KR, Miller DR, Reaven PD, Lynch JA et al. 2022. Using Mendelian randomisation to identify opportunities for type 2 diabetes prevention by repurposing medications used for lipid management. *EBioMedicine* **80**: 104038.

King EA, Davis JW, Degner JF. 2019. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *PLoS Genet* **15**: e1008489.

Kumar V, Krishna S, Siddiqi MI. 2015. Virtual screening strategies: recent advances in the identification and design of anti-cancer agents. *Methods* **71**: 64-70.

MacEachern SJ, Forkert ND. 2021. Machine learning for precision medicine. *Genome* **64**: 416-425.

Mirza SB, Salmas RE, Fatmi MQ, Durdagi S. 2016. Virtual screening of eighteen million compounds against dengue virus: Combined molecular docking and molecular dynamics simulations study. *J Mol Graph Model* **66**: 99-107.

Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, Floratos A, Sham PC, Li MJ, Wang J et al. 2015. The support of human genetic evidence for approved drug indications. *Nat Genet* **47**: 856-860.

Nielsch U, Fuhrmann U, Jaroch S. 2016. New Approaches to Drug Discovery. In *Handbook of Experimental Pharmacology*,, doi:10.1007/978-3-319-28914-4, pp. 1 online resource (VIII, 341 pages 100 illustrations, 363 illustrations in color. Springer International Publishing : Imprint: Springer,, Cham.

Shuey MM, Lee KM, Keaton J, Khankari NK, Breeyear JH, Walker VM, Miller DR, Heberer KR, Reaven PD, Clarke SL et al. 2022. A genetically supported drug repurposing pipeline for diabetes treatment using electronic health records. *medRxiv* doi:10.1101/2022.12.14.22283414: 2022.2012.2014.22283414.

Wang Y, Lin HQ, Wang P, Hu JS, Ip TM, Yang LM, Zheng YT, Chi-Cheong Wan D. 2017. Discovery of a Novel HIV-1 Integrase/p75 Interacting Inhibitor by Docking Screening, Biochemical Assay, and in Vitro Studies. *J Chem Inf Model* **57**: 2336-2343.

Wang Y, Yang Y, Chen S, Wang J. 2021. DeepDRK: a deep learning framework for drug repurposing through kernel-based multi-omics integration. *Brief Bioinform* **22**.

Wouters OJ, McKee M, Luyten J. 2020. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. *JAMA* **323**: 844-853.

Xu J, Mao C, Hou Y, Luo Y, Binder JL, Zhou Y, Bekris LM, Shin J, Hu M, Wang F et al. 2022. Interpretable deep learning translation of GWAS and multi-omics findings to identify pathobiology and drug repurposing in Alzheimer's disease. *Cell Rep* **41**: 111717.